DOCUMENT RESUME

ED 296 696 IR 013 268

AUTHOR Flagg, Barbara N.
TITLE Formative Evaluation in the Development Phases of a College Telecourse.
PUB DATE [Oct 87]
NOTE 18p.; Paper presented at the Annual Meeting of the American Evaluation Association (Boston, MA, October 15-17, 1987).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Educational Television; Evaluation Criteria; *Formative Evaluation; Higher Education; Instructional Effectiveness; *Material Development; *Mathematics Instruction; Pretests Posttests; Statistics; Teacher Characteristics; *Telecourses

ABSTRACT

This paper summarizes formative evaluation activities during the development of a pilot program by the Consortium for Mathematics and Its Applications for a telecourse on introductory data analysis. These activities included: (1) two major design activities--selection of a statistician host teacher and script development; (2) production phase activities--testing of the rough-cut pilot with students and statistics faculty in order to assess the appeal, comprehensibility, and recall of the materials; and (3) implementation activities--viewing of the pilot unit by students in statistics classes across the country, who also read the associated textbook materials and did the appropriate homework exercises. (The formative evaluation plan for the rough-cut pilot has been replicated for the final version of the television program.) Pretest-posttest performance by 300 students demonstrated a significant increase in achievement of the learning objectives. Included are six figures showing evaluation questions, findings, and actions taken for the three phases as well as data on recall, appeal, and comprehensibility. (MES)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED296696

American Evaluation Association

Formative Evaluation in the Development Phases of a College Telecourse

Barbara N. Flagg, Harvard University

Paper presented at the meeting of the American Evaluation Association, Boston, MA, October, 1987

Introduction

The Consortium for Mathematics and Its Applications (COMAP) is developing a telecourse for college students on introductory statistics. The development is sponsored by the Annenberg/Corporation for Public Broadcasting Project.

A telecourse is a college credit semester course that has a textbook and homework assignments but uses television in place of the lectures. Students who sign up with their local college as distant learners will do their work at home and meet occasionally with a professor on campus.

Over the past six months the COMAP project team has developed a pilot learning unit consisting of one half-hour television program and associated textbook and homework materials. This presentation will summarize the formative evaluation activities during the design, production, and implementation phases of the pilot program for the statistics telecourse.

IR013268

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara N. Flagg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Learning Unit Effectiveness

I would like to begin by telling the end of the story first. Two weeks ago, 300 students in Introductory Statistics courses across the country viewed the pilot television program, read the textbook material and did the homework assignment. The pilot learning unit discussed linear and exponential models for growth.

When pretest performance was compared with posttest performance on criterion-referenced questions, students demonstrated a significant increase in achievement of the learning objectives ($p < .01$). Thus, the one-group pretest-posttest evaluation gave evidence as to the instructional effectiveness of the pilot learning unit.

Leading up to this successful ending has been six months of formative evaluation activities during the design and production phases of the pilot television program. The production staff used information from each formative study to improve the potential effectiveness of the program materials.

Design Phase and Pre-production Formative Evaluation

During the Design Phase, there were two major pre-production formative evaluation activities. One effort involved the choice of a program host, while the second concerned script development.

Host teacher studies. Because the proposed format of the television program called for a statistician host teacher, one of the first decisions of the Design Phase was who to choose for the host teacher.

Insert Figure 1 about here

The goals of the pre-production formative evaluation, therefore, were to determine the appeal of a variety of host teachers and to explore what qualities supported or detracted from that appeal. Not only did the producers desire students' feedback as to which host teacher to choose but they also wanted information that would help them direct the host teacher in the pilot production.

Two host teacher studies were carried out with a total of 95 students from two- and four-year institutions in Boston. The students viewed videotapes of auditions of local statisticians. Each host teacher candidate read a short piece of script and then using provided props presented an extemporaneous lesson on the making of and meaning of an histogram.

After viewing each host teacher, students gave a one word description of their immediate feeling about the candidate and then rated the candidate's appeal, enthusiasm, approachability, and confidence [illegible] ts

also comparatively ranked the candidates and provided reasons for their first choice host teacher.

Participating students expressed a wide diversity of opinions about who was the best host teacher. The data analysis pointed to three characteristics which appeared to determine students' choices: an approachable personal manner; an attractive physical appearance, and a pleasant voice quality. However, an approachable personal manner was found to detract from appeal when it was too casual or "laid-back."

Although none of the candidates won a clear majority of first choice votes, one statistician (let's call her Teresa) did appeal consistently across the whole range of students. Teresa received high scores on all the rated characteristics and received the majority of positive student comments with very few negative comments. She was portrayed by students as "very warm and pleasant" with a "good speaking voice" and a "clear and understandable" presentation.

None of the auditioning statisticians turned out to be the charismatic television personality for whom the producers were looking. However, the formative evaluation data confirmed the producers' intuitions that Teresa was "a pleaser who would provoke very little negative reaction." The diversity of the telecourse target audience required someone with mass audience appeal; so Teresa was asked to host the television series.

Script development. Another activity of the Design Phase was to develop an effective pilot script.

Insert Figure 2 about here

The goals of the pre-production formative evaluation at this time were to obtain feedback as to the clarity, accuracy, and comprehensiveness of the content and to assess the strengths and weaknesses of the program's instructional format. The format of the program involved documentaries and animated graphics that demonstrated the application of statistical concepts to real-life problems. The format also used a host statistician who introduced, elaborated, and summarized the statistical concepts.

This past June, a draft pilot script was reviewed by five Statistics professors from a variety of institutions and states. There was astonishing agreement among the reviewers about what was strong and what was weak in the script, as well as about specific changes. The general strengths applauded by the reviewers included the examples chosen for the documentaries and the level of explanation in the script. The weaknesses identified included statistical inaccuracies and terminology problems, as well as a feeling that the human interest aspects of the documentary examples were given precedence over the

statistical points. Moreover, the statisticians believed that the content objectives were not obvious.

In response to the expert review study, the production staff rewrote the script, correcting inaccuracies and tightening terminology. The content objectives were more clearly defined in the introductory and concluding segments. In the documentary segments, the narrator's explanation of statistical concepts was expanded. And finally, the amount of material which was entertaining but irrelevant to the statistical points was decreased. The production staff's goal in making these changes was to achieve a balance between the educational objectives and the motivating audiovisual quality of television.

Production Phase and Production Formative Evaluation

Once the script was reworked to reflect the feedback of the subject matter experts, the project team moved into a production phase.

Insert Figure 3 about here

Initial efforts focused on producing an unpolished videotape version of the first half of the revised script.

The topic was the linear model of growth. The rough cut program began with a brief overview by a female host (Teresa was not available at this time). Two documentaries followed describing the use of growth curves to assess the need for growth hormone therapy in slow growing children. Finally, the host introduced concepts including the least squares line, the pattern of residuals, and prediction.

Rough cut study. The rough cut pilot was tested with the target student audience and statistics faculty to provide feedback to the production staff about their format decisions. The major goals of the production formative evaluation were to assess the appeal and comprehensibility of program segments as well as to measure student short-term recall.

An 18 minute videotape was shown to 53 students in Introductory Statistics classes in 3 states.

Insert Figure 4 about here

Student recall of the program content was measured with four recall questions. These questions were answered by a random half of the students prior to viewing and by

all of the students after viewing. Viewing the program resulted in a significant increase ($p < .01$) in knowledge on three out of the four questions.

In addition to measuring recall outcomes, the formative evaluation attempted to analyze what was appealing and understandable in the program and what was not.

Insert Figure 5 about here

To this end, a continuous reaction measure was employed. While viewing the rough cut program, half of the students rated the appeal of the material at 7 different points. The other half of the students rated the comprehensibility of the material at the same 7 points. Students were randomly assigned to rate either appeal or comprehensibility. The researcher called out the numbers 1-7 to indicate when students were to rate the material while the videotape continued running.

Students rated the majority of the program as understandable but appeal remained high only during the documentary segments, dropping off when the host teacher explained some additional statistical concepts.

In addition to the continuous reaction measure, student response was elicited about the program pace, the style of graphics, what parts they liked most and least, and what concepts they felt needed more explanation (see Figure 3). Generally, students rated the pace of statistical concepts as too fast and were only midly positive about the informal graphic style used in the program.

Ratings on the continuous reaction measures, posttest recall results, and student and faculty responses to postviewing questions complement each other to give the project team a picture of what works in the program, what doesn't work, and possible reasons why. For example, the continuous reaction measure showed a dip in comprehensibility and appeal at point #6 (see Figure 5). At this point in the program, the host teacher is narrating a series of graphic displays about the pattern of residuals or deviations from the least squares line. The pretest results showed that this was a new concept to the students, and the posttest results indicated a significant improvement in understanding this concept, but only 42% of the sample gave the correct posttest answer (see Figure 4). When asked what topics in the program needed further explanation, 52% of the viewers wanted more about the pattern of residuals. The students did not feel confident of their knowledge of this concept. Repeated

confirmation of a difficulty in the program through differing measuring instruments can convince producers that serious revision is necessary; just adding a music track won't work. Inductive analysis of the complementary data sets provided the production staff with information on which to base revisions in the television program format.

The production staff concluded that the documentaries could handle more graphics and discussion about the statistical objectives without sacrificing audience interest and understanding (see Figure 3). They decided to decrease the host on-air exposure, while carrying one or more of the content objectives with narrated graphics. Further, each program would be limited to 5-7 content objectives in order to treat them in more depth at a slower pace. Finally, the graphic style would change to a slicker commercial television look.

Implementation Phase and Implementation Formative Evaluation

We are currently analyzing data from the final pilot phase, that is, from the implementation of the pilot learning unit in the classroom (see Figure 6).

Insert Figure 6 about here

The formative evaluation plan for the rough cut pilot has been replicated for the final television program. Additionally, the complete learning unit was implemented in statistics classrooms in 4 states. The goals of this formative evaluation are to detemine whether students achieve the learning objectives, how confident they feel about their knowledge, and how they feel about the curriculum components separately and together.

As with the previous studies described, the implementation formative evaluation allows the student and faculty users to have a voice in the design of the statistics telecourse. The producers will consider the feedback from students and faculty about the strengths and weaknesses of the pilot learning unit when developing plans for the next 25 statistics telecourse units.

In summary, formative evaluation has played a critical role in helping to guide the design and revisions of the pilot learning unit of COMAP's statistics telecourse.

Figure 1

DESIGN PHASE ACTIVITY: HOST TEACHER CHOICE

PRE-PRODUCTION FORMATIVE EVALUATION:

EVALUATION QUESTIONS

WHICH HOST-TEACHER DO STUDENTS FIND APPEALING?

WHAT QUALITIES SUPPORT OR DETRACT FROM APPEAL?

FINDINGS

QUALITIES SUPPORTING APPEAL: APPROACHABLE PERSONAL MANNER

ATTRACTIVE PHYSICAL APPEARANCE

PLEASANT VOICE QUALITY

QUALITIES DETRACTING FROM APPEAL: CASUAL/LAID-BACK PERSONAL MANNER

ACTION TAKEN

"TERESA" CHOSEN AS HOST

Figure 2

DESIGN PHASE: PILOT SCRIPT DEVELOPMENT

PRE-PRODUCTION FORMATIVE EVALUATION:

EVALUATION QUESTIONS

IS THE SCRIPT CLEAR? ACCURATE? COMPREHENSIVE?

WHAT ARE THE PROGRAM FORMAT'S STRENGTHS AND WEAKNESSES?

FINDINGS

STRENGTHS: GOOD EXAMPLES FOR DOCUMENTARIES

APPROPRIATE LEVEL OF EXPLANATION

WEAKNESSES: STATISTICAL INACCURACIES

LOOSE USE OF TERMINOLOGY

HUMAN INTEREST ASPECTS OF DOCUMENTARIES GIVEN PRECEDENCE OVER STATISTICAL POINTS

CONTENT OBJECTIVES UNCLEAR

ACTION TAKEN

CORRECTION OF INACCURACIES AND TERMINOLOGY

CLEARER OPENING AND CLOSING STATEMENTS OF CONTENT OBJECTIVES

INCREASE IN AMOUNT OF EXPLANATION OF STATISTICAL CONCEPTS BY NARRATOR IN DOCUMENTARIES

DECREASE IN AMOUNT OF ENTERTAINING BUT CONTENT-IRRELEVANT MATERIAL

Figure 3

PRODUCTION PHASE: ROUGH CUT PILOT

PRODUCTION FORMATIVE EVALUATION:

EVALUATION QUESTIONS

WHAT PARTS DID STUDENTS FIND APPEALING?

WHAT PARTS DID STUDENTS FIND COMPREHENSIBLE?

DID STUDENTS RECALL PROGRAM CONTENT?

FINDINGS

SIGNIFICANT IMPROVEMENT ON 3 OUT OF 4 CONTENT QUESTIONS

DOCUMENTARIES APPEALING AND UNDERSTANDABLE

HOST TEACHER SEGMENTS LESS APPEALING AND LESS UNDERSTANDABLE

PACE OF STATISTICAL CONCEPTS TOO FAST

GRAPHIC STYLE MILDLY APPEALING

ACTION TAKEN

INCREASE GRAPHICS AND DISCUSSION ABOUT STATISTICAL OBJECTIVES IN DOCUMENTARIES

DECREASE HOST ON-AIR TIME

USE NARRATED GRAPHICS TO CONVEY ONE (OR MORE) OF OBJECTIVES

LIMIT EACH PROGRAM TO 5-7 CONTENT OBJECTIVES

INCREASE DYNAMICS AND PROFESSIONAL TV LOOK OF GRAPHICS

Figure 4

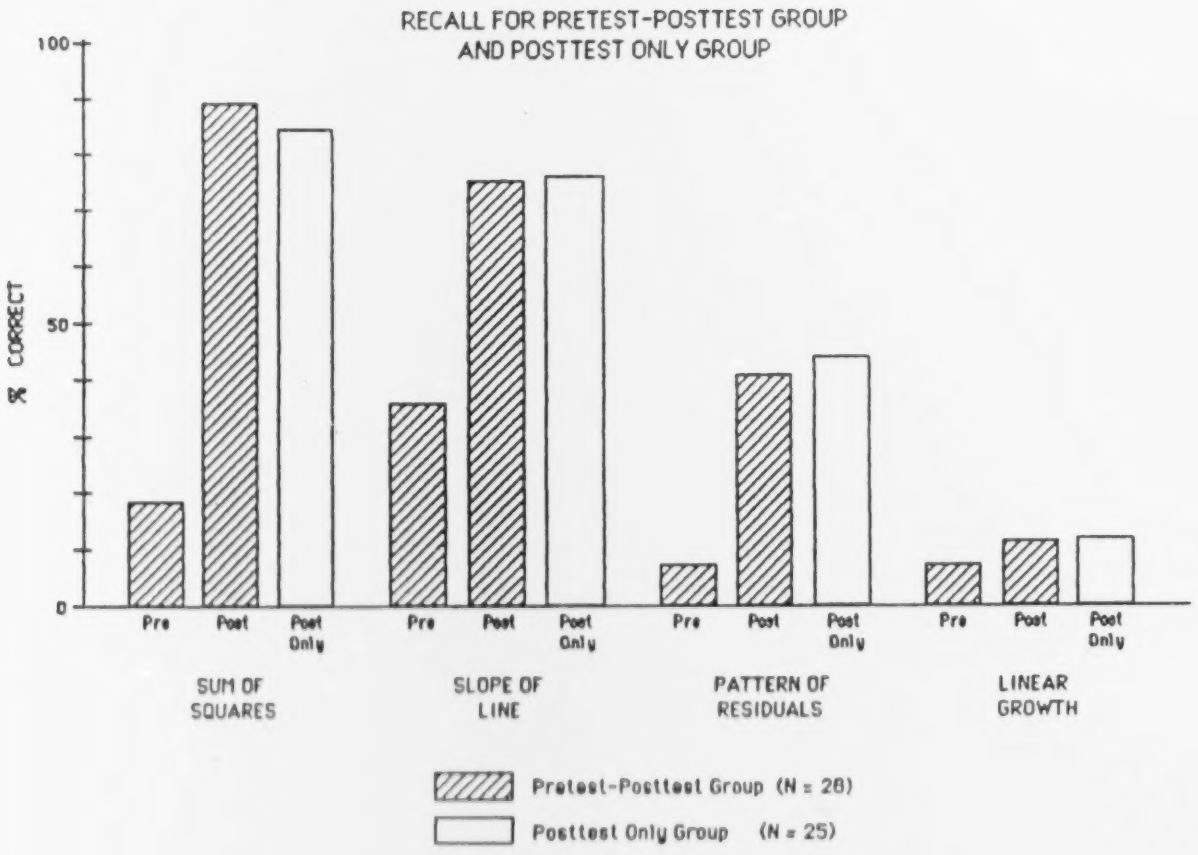
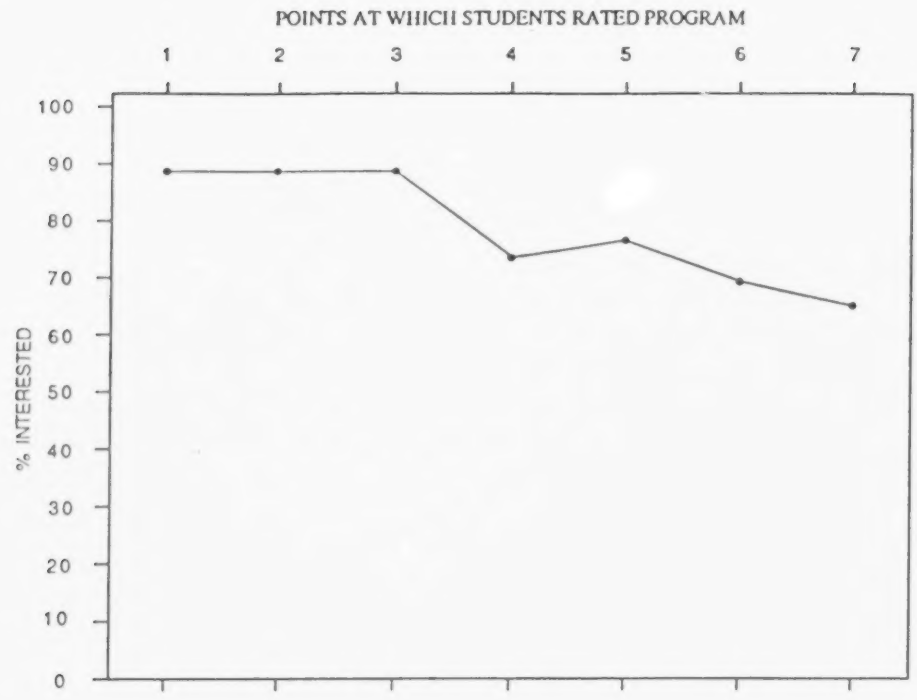

RECALL FOR PRETEST-POSTTEST GROUP
AND POSTTEST ONLY GROUP
100
50
0
% CORRECT
Pre
Post
Post Only
SUM OF SQUARES
SLOPE OF LINE
PATTERN OF RESIDUALS
LINEAR GROWTH
Pretest-Posttest Group (N = 28)
Posttest Only Group (N = 25)

Figure 5

APPEAL RATINGS (N = 26)



COMPREHENSIBILITY RATINGS (N = 27)

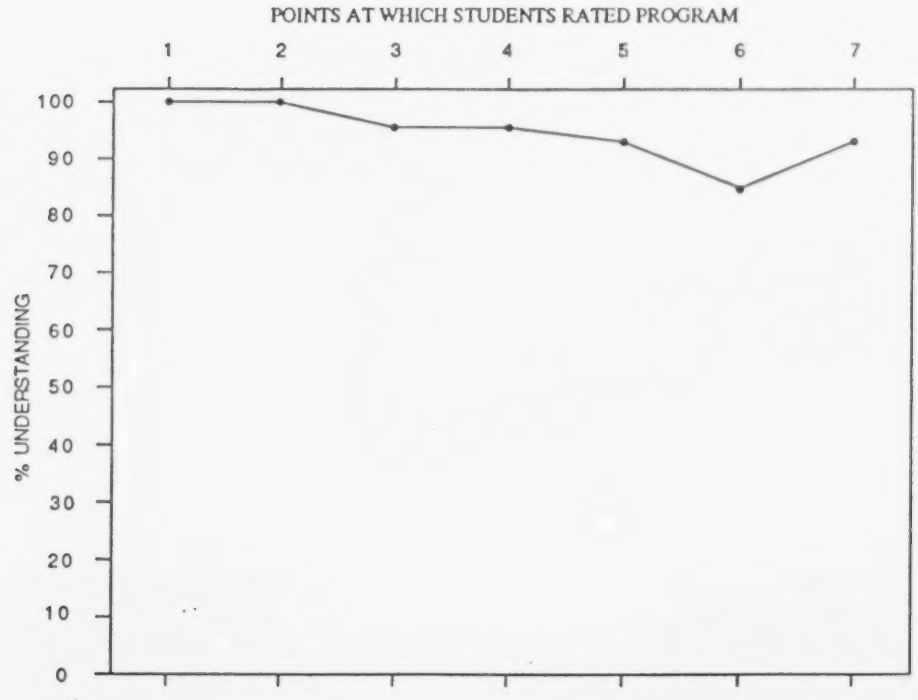


Figure 6

IMPLEMENTATION PHASE: PILOT LEARNING UNIT COMPRISES HALF-HOUR TELEVISION PROGRAM AND TEXTBOOK WITH HOMEWORK

IMPLEMENTATION FORMATIVE EVALUATION:

EVALUATION QUESTIONS FOR TV PROGRAM

WHAT PARTS DID STUDENTS FIND APPEALING?

WHAT PARTS DID STUDENTS FIND COMPREHENSIBLE?

DID STUDENTS RECALL PROGRAM CONTENT?

EVALUATION QUESTIONS FOR LEARNING UNIT

DID STUDENTS ACHIEVE THE LEARNING OBJECTIVES?

HOW CONFIDENT DID STUDENTS FEEL ABOUT THEIR KNOWLEDGE?

HOW DID STUDENTS RATE THE TV PROGRAM, TEXT, AND EXERCISES SEPARATELY ON DIFFICULTY, CLARITY, COMPLETENESS, HELPFULNESS, AND APPEAL?

HOW DID STUDENTS RATE THE UNIT AS A WHOLE?